

Selected examples for Chapter 6.

EXAMPLE 6.1

A constant but unknown signal is observed in additive Gaussian white noise. That is

$$x[n] = A + w[n]$$

where A is the unknown signal, $w[n]$ is the noise, and $x[n]$ is the observed random process. If A and $w[n]$ are complex, then the noise at any point n has the complex Gaussian density function

$$f_w(w) = \frac{1}{\pi\sigma_w^2} e^{-\frac{|w|^2}{\sigma_w^2}}$$

If \mathbf{x} represents the vector of samples $x[0], x[1], \dots, x[N_s - 1]$, then the observations for this example are independent and the joint density for the observations is a product of marginal densities. The likelihood function is therefore

$$f_{\mathbf{x};A}(\mathbf{x}; A) = \prod_{n=0}^{N_s-1} \frac{1}{\pi\sigma_w^2} e^{-\frac{|x[n]-A|^2}{\sigma_w^2}}$$

and the log likelihood function is

$$\ln f_{\mathbf{x};A}(\mathbf{x}; A) = -N_s \ln(\pi\sigma_w^2) - \sum_{n=0}^{N_s-1} \frac{|x[n] - A|^2}{\sigma_w^2}$$

Instead of rearranging this to show the explicit dependence on A we can simply observe that the log likelihood function is continuous and set the complex gradient to zero. Applying the formulas in Table A.3 yields

$$\nabla_A \ln f_{\mathbf{x};A}(\mathbf{x}; A) = - \sum_{n=0}^{N_s-1} \frac{(x[n] - A)^*}{\sigma_w^2} = 0$$

Solving this for A produces the maximum likelihood estimate for the signal:

$$\hat{A}_{ml} = \frac{1}{N_s} \sum_{n=0}^{N_s-1} x[n]$$

This is the sample mean of the random process. □

EXAMPLE 6.3

Assume that the set of observations in an experiment x_1, x_2, \dots, x_N are independent and that the random variables x_i each have mean m and variance σ_o^2 . The sample mean is given by

$$\hat{m} = \frac{1}{N} \sum_{i=1}^N x_i$$

The mean of this estimate is

$$\mathcal{E}\{\hat{m}\} = \frac{1}{N} \sum_{i=1}^N \mathcal{E}\{x_i\} = \frac{1}{N} N m = m$$

Thus the sample mean is unbiased. The variance of the estimate is given by

$$\text{Var}[\hat{m}] = \frac{1}{N^2} \sum_{i=1}^N \text{Var}[x_i] = \frac{1}{N^2} N \sigma_o^2 = \frac{\sigma_o^2}{N}.$$

Since the estimate is unbiased and the variance decreases with N , the estimate is consistent. This implies that as the number of samples gets very large, the probability that the estimate differs from the *true* value of the mean approaches zero. □

EXAMPLE 6.4

Consider a set of observations $\mathbf{x} = [x_1 x_2 \cdots x_N]^T$ from the one-dimensional Gaussian density function

$$f_{x;m}(\mathbf{x}; m) = \frac{1}{\sqrt{2\pi\sigma_o^2}} e^{-\frac{(\mathbf{x}-m)^2}{2\sigma_o^2}}$$

The log likelihood function for the mean is given by

$$\ln f_{\mathbf{x};m}(\mathbf{x}; m) = -N \ln(\sqrt{2\pi\sigma_o^2}) - \sum_{i=1}^N \frac{(x_i - m)^2}{2\sigma_o^2}$$

The derivative of the log likelihood function is

$$\frac{\partial \ln f_{\mathbf{x};m}(\mathbf{x}; m)}{\partial m} = \sum_{i=1}^N \frac{(x_i - m)}{\sigma_o^2}$$

Since the x_i are uncorrelated, the cross terms are zero and the expectation of this squared quantity is

$$\mathcal{E} \left\{ \left(\frac{\partial \ln f_{\mathbf{x};m}(\mathbf{x}; m)}{\partial m} \right)^2 \right\} = \sum_{i=1}^N \frac{\mathcal{E} \{ (x_i - m)^2 \}}{\sigma_o^4} = \frac{N\sigma_o^2}{\sigma_o^4} = \frac{N}{\sigma_o^2}$$

Therefore, by the Cramér-Rao inequality

$$\text{Var} [\hat{\theta}] \geq \frac{1}{\mathcal{E} \left\{ \left(\frac{\partial \ln f_{\mathbf{x};\theta}(\mathbf{x};\theta)}{\partial \theta} \right)^2 \right\}}$$

the bound on the variance of *any* unbiased estimate for the mean of the Gaussian density is

$$\text{Var} [\hat{m}] \geq \frac{\sigma_o^2}{N}$$

This result can also be achieved using

$$\text{Var} [\hat{\theta}] \geq \frac{1}{-\mathcal{E} \left\{ \frac{\partial^2 \ln f_{\mathbf{x};\theta}(\mathbf{x};\theta)}{\partial \theta^2} \right\}}$$

since

$$-\frac{\partial^2 \ln f_{\mathbf{x};m}(\mathbf{x};m)}{\partial m^2} = -\frac{\partial}{\partial m} \left(\sum_{i=1}^N \frac{(x_i - m)}{\sigma_o^2} \right) = \sum_{i=1}^N \frac{1}{\sigma_o^2} = \frac{N}{\sigma_o^2}$$

Note from the variance expression in Example 6.3 that the sample mean satisfies this bound with equality. This result is not really surprising, since it was shown in Section 6.11 of the text that the sample mean is the maximum likelihood estimate for the mean of the Gaussian density.

To show that the sample mean satisfies the condition necessary for a minimum variance estimate, note from above that the derivative of the log likelihood function can be expressed as

$$\frac{\partial \ln f_{\mathbf{x};m}(\mathbf{x}; m)}{\partial m} = \frac{1}{\sigma_o^2} \sum_{i=1}^N (x_i - m) = \frac{N}{\sigma_o^2} \left(\left[\frac{1}{N} \sum_{i=1}^N x_i \right] - m \right)$$

or

$$(\hat{m}_{ml} - m) = \frac{\sigma_o^2}{N} \cdot \frac{\partial \ln f_{\mathbf{x};m}(\mathbf{x}; m)}{\partial m}$$

where \hat{m}_{ml} is the sample mean. This last equation is in the form of

$$\hat{\theta}(\mathbf{x}) - \theta = K(\theta) \cdot \frac{\partial \ln f_{\mathbf{x};\theta}(\mathbf{x}; \theta)}{\partial \theta}$$

□

EXAMPLE 6.5

A d -dimensional real random vector \mathbf{v} is described by a Gaussian density function

$$f_{\mathbf{v}}(\mathbf{v}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\mathbf{C}_{\mathbf{v}}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{v}-\mathbf{m})^T \mathbf{C}_{\mathbf{v}}^{-1}(\mathbf{v}-\mathbf{m})}$$

Given N independent samples of the random vector $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(N)}$ it is desired to form a maximum likelihood estimate for the mean vector \mathbf{m} .

The likelihood function for this problem is

$$f_{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(N)}; \mathbf{m}}(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(N)}; \mathbf{m}) = \prod_{i=1}^N \frac{1}{(2\pi)^{\frac{d}{2}} |\mathbf{C}_{\mathbf{v}}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{v}^{(i)} - \mathbf{m})^T \mathbf{C}_{\mathbf{v}}^{-1}(\mathbf{v}^{(i)} - \mathbf{m})}$$

The log likelihood function is thus

$$\ln f_{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(N)}; \mathbf{m}}(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(N)}; \mathbf{m}) = \text{const.} - \frac{1}{2} \sum_{i=1}^N (\mathbf{v}^{(i)} - \mathbf{m})^T \mathbf{C}_{\mathbf{v}}^{-1}(\mathbf{v}^{(i)} - \mathbf{m})$$

Taking the gradient with respect to \mathbf{m} and setting it to zero yields

$$\nabla_{\mathbf{m}} \ln f_{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(N)}; \mathbf{m}}(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(N)}; \mathbf{m}) = \sum_{i=1}^N \mathbf{C}_{\mathbf{v}}^{-1}(\mathbf{v}^{(i)} - \mathbf{m}) = \mathbf{0}$$

or, since $\mathbf{C}_{\mathbf{v}}^{-1}$ is a constant,

$$\hat{\mathbf{m}}_N = \frac{1}{N} \sum_{i=1}^N \mathbf{v}^{(i)}$$

This is the sample mean vector.

Let us check the properties of this estimate. The expected value is

$$\mathcal{E}\{\hat{\mathbf{m}}_N\} = \frac{1}{N} \sum_{i=1}^N \mathcal{E}\{\mathbf{v}^{(i)}\} = \frac{1}{N}(N\mathbf{m}) = \mathbf{m}$$

so the estimate is unbiased. The covariance matrix of the estimate is

$$\mathcal{E}\{(\hat{\mathbf{m}}_N - \mathbf{m})(\hat{\mathbf{m}}_N - \mathbf{m})^T\} = \frac{1}{N^2} \mathcal{E}\left\{\left(\sum_{i=1}^N (\mathbf{v}^{(i)} - \mathbf{m})\right) \left(\sum_{j=1}^N (\mathbf{v}^{(j)} - \mathbf{m})^T\right)\right\}$$

Since the $\mathbf{v}^{(i)}$ are independent, this reduces to

$$\frac{1}{N^2} \sum_{i=1}^N \mathcal{E}\{(\mathbf{v}^{(i)} - \mathbf{m})(\mathbf{v}^{(i)} - \mathbf{m})^T\} = \frac{1}{N} \mathbf{C}_{\mathbf{v}}$$

Since the estimate is unbiased and the covariance of the estimate decreases with N , the estimate is consistent.

Finally, let us check the Cramér Rao bound. The foregoing analysis shows

$$\mathbf{s} = \nabla_{\mathbf{m}} \ln f_{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(N)}; \mathbf{m}}(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(N)}; \mathbf{m}) = \sum_{i=1}^N \mathbf{C}_{\mathbf{v}}^{-1}(\mathbf{v}^{(i)} - \mathbf{m})$$

Since this can be put in the form

$$\mathbf{s} = \sum_{i=1}^N \mathbf{C}_{\mathbf{v}}^{-1}(\mathbf{v}^{(i)} - \mathbf{m}) = N \mathbf{C}_{\mathbf{v}}^{-1} \left(\frac{1}{N} \left(\sum_{i=1}^N \mathbf{v}^{(i)} \right) - \mathbf{m} \right) \text{ or } \hat{\mathbf{m}}_N - \mathbf{m} = (1/N) \mathbf{C}_{\mathbf{v}} \mathbf{s}$$

$\hat{\mathbf{m}}_N$ is evidently a minimum-variance estimate. However, let us proceed to check the bound explicitly. The Fisher information matrix is

$$\begin{aligned} \mathbf{J} = \mathcal{E} \{ \mathbf{s} \mathbf{s}^T \} &= \mathcal{E} \left\{ \left(\sum_{i=1}^N \mathbf{C}_{\mathbf{v}}^{-1}(\mathbf{v}^{(i)} - \mathbf{m}) \right) \left(\sum_{j=1}^N \mathbf{C}_{\mathbf{v}}^{-1}(\mathbf{v}^{(j)} - \mathbf{m}) \right)^T \right\} \\ &= \mathbf{C}_{\mathbf{v}}^{-1} \sum_{i=1}^N \mathcal{E} \{ (\mathbf{v}^{(i)} - \mathbf{m})(\mathbf{v}^{(i)} - \mathbf{m})^T \} \mathbf{C}_{\mathbf{v}}^{-1} \\ &= \mathbf{C}_{\mathbf{v}}^{-1} (N \mathbf{C}_{\mathbf{v}}) \mathbf{C}_{\mathbf{v}}^{-1} = N \mathbf{C}_{\mathbf{v}}^{-1} \end{aligned}$$

Thus

$$\mathbf{J}^{-1} = \frac{1}{N} \mathbf{C}_{\mathbf{v}}$$

and the variance of the estimate and the Cramér Rao bound are identical. \square

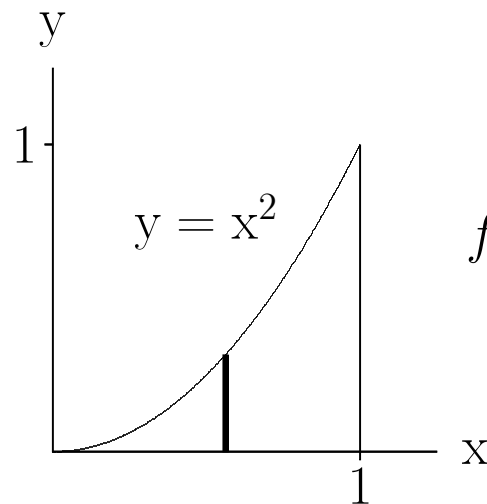
EXAMPLE 6.6

The power y in some unobserved signal is at most equal to the square of the magnitude x of a signal that *is* observed. The joint density function for the two random variables x and y is

$$f_{xy}(x, y) = \begin{cases} 10y & 0 \leq y \leq x^2, \quad 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

It is desired to estimate the power y from an observation x using both mean-square and MAP estimation.

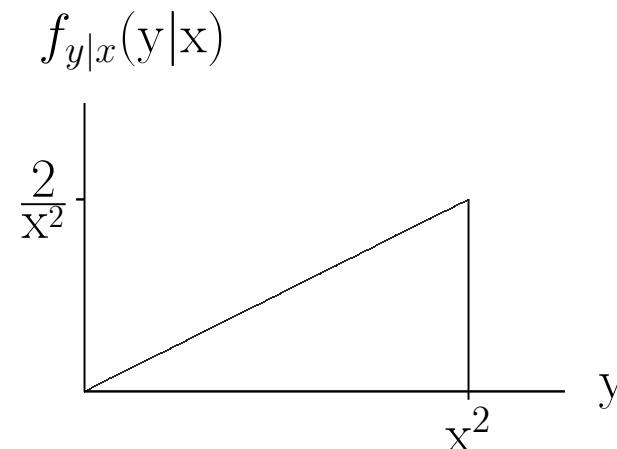
The marginal density for x can be computed by integrating as shown below



$$f_x(x) = \int_0^{x^2} 10y dy = 5y^2 \Big|_0^{x^2} = 5x^4 \quad 0 \leq x \leq 1$$

The conditional density for y is therefore

$$f_{y|x}(y|x) = \frac{10y}{5x^4} = \frac{2y}{x^4} \quad 0 \leq y \leq x^2$$



From this sketch it is clear that the maximum of the conditional density occurs at $y = x^2$. Therefore

$$\hat{y}_{MAP} = x^2$$

The mean of the conditional density is given by

$$\int_{-\infty}^{\infty} y f_{y|x}(y|x) dy = \int_0^{x^2} \frac{2y^2}{x^4} dy = \left. \frac{2y^3}{3x^4} \right|_0^{x^2} = \frac{2}{3}x^2$$

thus

$$\hat{y}_{ms} = \frac{2}{3}x^2$$

The minimum mean-square error is given by

$$\mathcal{E}\{(y - \hat{y}_{ms})^2\} = \int_0^1 \int_0^{x^2} (y - \tfrac{2}{3}x^2)^2 10y dy dx = \frac{5}{162} = 0.0309$$

To show that this mean-square error is less than say the mean-square error for the MAP estimate, compute

$$\mathcal{E}\{(y - \hat{y}_{MAP})^2\} = \int_0^1 \int_0^{x^2} (y - x^2)^2 10y dy dx = \frac{5}{54} = 0.0926$$

which turns out to be exactly three times the previous mean-square error. \square

EXAMPLE 6.7

The real random variables x and y have a joint Gaussian density with parameters

$$\mathbf{m}_{xy} = \begin{bmatrix} m_x \\ m_y \end{bmatrix} \quad \mathbf{C}_{xy} = \begin{bmatrix} \sigma_x^2 & c_{xy} \\ c_{xy} & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

where ρ_{xy} is the normalized correlation coefficient $c_{xy}/\sigma_x\sigma_y$. The joint density can be written as

$$\begin{aligned} f_{xy}(x, y) &= \frac{1}{(2\pi)|\mathbf{C}_{xy}|^{\frac{1}{2}}} e^{-\frac{1}{2}[\mathbf{x}-\mathbf{m}_{xy}]\mathbf{C}_{xy}^{-1}\begin{bmatrix} x-m_x \\ y-m_y \end{bmatrix}} \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} e^{-\frac{1}{2(1-\rho_{xy}^2)}\left\{\frac{(x-m_x)^2}{\sigma_x^2} - 2\rho_{xy}\frac{(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right\}} \end{aligned}$$

The marginal density for x can be found by integration to be

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}$$

Then by forming the ratio

$$f_{y|x}(y|x) = \frac{f_{xy}(x, y)}{f_x(x)}$$

and simplifying we obtain

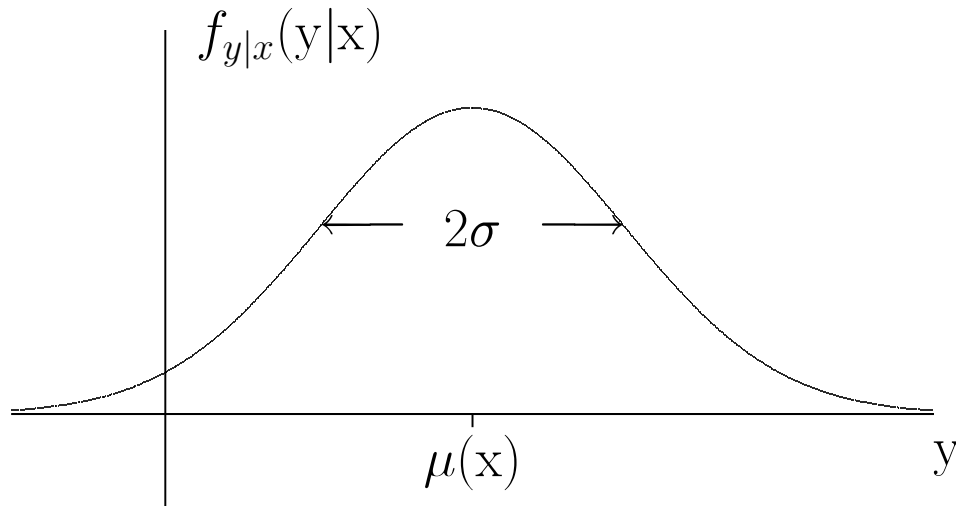
$$f_{y|x}(y|x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu(x))^2}{2\sigma^2}}$$

where

$$\mu(x) = m_y + \rho_{xy} \frac{\sigma_y}{\sigma_x} (x - m_x) = \left(\rho_{xy} \frac{\sigma_y}{\sigma_x} \right) x + \left(m_y - \rho_{xy} \frac{\sigma_y}{\sigma_x} m_x \right)$$

and

$$\sigma^2 = \sigma_y^2 (1 - \rho_{xy}^2)$$



Since the conditional density is Gaussian, *the mean and the maximum occur at the same place*, consequently

$$\hat{y}_{ms} = \hat{y}_{MAP} = \mu(x) = \left(\rho_{xy} \frac{\sigma_y}{\sigma_x} \right) x + \left(m_y - \rho_{xy} \frac{\sigma_y}{\sigma_x} m_x \right)$$

Note that the function $\mu(x)$ plots as a straight line. It is common in this case to say that the estimate is a *linear* function of x although, strictly speaking $\mu(x)$ is not a linear function. Let us instead say that \hat{y} has a *linear dependence* on x . A repetition of this analysis for the complex case shows that the last formula also applies when x and y are complex.

The mean-square error corresponding to this estimate can be shown to be equal to the variance σ^2 of the conditional density (see Problem 6.17). This is also the lower bound given by

$$\mathcal{E} \{ (y - \hat{y})^2 \} \geq \frac{1}{\mathcal{E} \left\{ \left(\frac{\partial \ln f_y \mathbf{x}(y, \mathbf{x})}{\partial y} \right)^2 \right\}} = \frac{1}{-\mathcal{E} \left\{ \frac{\partial^2 \ln f_y \mathbf{x}(y, \mathbf{x})}{\partial y^2} \right\}}$$

To show this it is easiest to use the second form on the right.

The logarithm of the joint density is

$$\ln f_{xy}(x, y) = \text{const.} - \frac{1}{2(1 - \rho_{xy}^2)} \left\{ \frac{(x - m_x)^2}{\sigma_x^2} - 2\rho_{xy} \frac{(x - m_x)(y - m_y)}{\sigma_x \sigma_y} + \frac{(y - m_y)^2}{\sigma_y^2} \right\}$$

Thus

$$\frac{\partial \ln f_{xy}(x, y)}{\partial y} = -\frac{1}{2(1 - \rho_{xy}^2)} \left\{ -2\rho_{xy} \frac{(x - m_x)}{\sigma_x \sigma_y} + \frac{2(y - m_y)}{\sigma_y^2} \right\}$$

and

$$\frac{\partial^2 \ln f_{xy}(x, y)}{\partial y^2} = -\frac{1}{2(1 - \rho_{xy}^2)} \left\{ \frac{2}{\sigma_y^2} \right\} = -\frac{1}{\sigma_y^2(1 - \rho_{xy}^2)} = -\frac{1}{\sigma^2}$$

Substituting this in (6.90) shows that the mean-square error for *any* estimate \hat{y} is lower bounded by

$$\mathcal{E}\{(y - \hat{y})^2\} \geq \sigma^2$$

□

EXAMPLE 6.9

The random variables in Example 6.6 have the joint density function

$$f_{xy}(x, y) = \begin{cases} 10y & 0 \leq y \leq x^2, \quad 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Here the linear mean-square estimate of y given x is computed and compared to the optimal (nonlinear) mean-square estimate computed in that example.

The needed first and second moment parameters are computed as follows:

$$m_x = \int_0^1 \int_0^{x^2} x \cdot 10y dy dx = \frac{5}{6} \quad m_y = \int_0^1 \int_0^{x^2} 10y^2 dy dx = \frac{10}{21}$$

$$\mathcal{E}\{x^2\} = \int_0^1 \int_0^{x^2} 10x^2y dy dx = \frac{5}{7} \quad \sigma_x^2 = \mathcal{E}\{x^2\} - m_x^2 = \frac{5}{7} - \left(\frac{5}{6}\right)^2 = \frac{5}{252}$$

$$\mathcal{E}\{y^2\} = \int_0^1 \int_0^{x^2} 10y^3 dy dx = \frac{5}{18} \quad \sigma_y^2 = \mathcal{E}\{y^2\} - m_y^2 = \frac{5}{18} - \left(\frac{10}{21}\right)^2 = \frac{5}{98}$$

$$\mathcal{E}\{xy\} = \int_0^1 \int_0^{x^2} 10xy^2 dy dx = \frac{5}{12}$$

$$c_{xy} = \mathcal{E}\{xy\} - m_x m_y = \frac{5}{12} - \left(\frac{5}{6}\right) \left(\frac{10}{21}\right) = \frac{5}{252}$$

$$\rho_{xy} = \frac{c_{xy}}{\sigma_x \sigma_y} = \sqrt{\frac{(5/252)^2}{(5/252)(5/98)}} = \sqrt{\frac{7}{18}}$$

Now we can compute

$$a = \frac{c_{xy}}{\sigma_x^2} = \frac{5/252}{5/252} = 1 \quad \text{and} \quad b = m_y - am_x = \frac{10}{21} - \frac{5}{6} = -\frac{5}{14}$$

Therefore the linear mean-square estimate is

$$\hat{y}_{lms} = ax + b = x - \frac{5}{14}$$

The corresponding mean-square error is

$$\mathcal{E}_{lms} = \sigma_y^2(1 - \rho_{xy}^2) = \frac{5}{98} \left(1 - \frac{7}{18}\right) = \frac{55}{1764} = 0.0312$$

A comparison to the results of Example 6.6, shows that this is only slightly worse than the mean-square error for the optimal (nonlinear) mean-square estimate, which produced $\mathcal{E}_{lms} = 0.0309$. \square

EXAMPLE 6.10

Two real random variables x_1 and x_2 and a related random variable y are jointly distributed. It is known that if \mathbf{v} is defined by

$$\mathbf{v} = \begin{bmatrix} y \\ x_1 \\ x_2 \end{bmatrix}$$

then the mean vector and covariance matrix of \mathbf{v} are

$$\mathbf{m}_\mathbf{v} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \quad \mathbf{C}_\mathbf{v} = \begin{bmatrix} \frac{7}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{10} & \frac{3}{10} & -\frac{1}{10} \\ \frac{1}{10} & -\frac{1}{10} & \frac{3}{10} \end{bmatrix}$$

It is desired to find the best linear mean-square estimate of y using x_1 and x_2 .

From the given information it follows that

$$\mathbf{m}_\mathbf{x} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \quad m_y = \frac{1}{4}$$
$$\mathbf{C}_\mathbf{x} = \begin{bmatrix} \frac{3}{10} & -\frac{1}{10} \\ -\frac{1}{10} & \frac{3}{10} \end{bmatrix} \quad \mathbf{c}_{\mathbf{x}y} = \begin{bmatrix} \frac{1}{10} \\ \frac{1}{10} \end{bmatrix} \quad \sigma_y^2 = \frac{7}{10}$$

Therefore to find \mathbf{a} form:

$$\begin{bmatrix} \frac{3}{10} & -\frac{1}{10} \\ -\frac{1}{10} & \frac{3}{10} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{10} \\ \frac{1}{10} \end{bmatrix}$$

and solve this to obtain

$$a_1 = a_2 = \frac{1}{2}$$

The constant b is then given by

$$b = m_y - \mathbf{a}^{*T} \mathbf{m}_x = \frac{1}{4} - \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = -\frac{1}{4}$$

so the optimal linear mean-square estimate is

$$\hat{y} = \frac{1}{2}x_1 + \frac{1}{2}x_2 - \frac{1}{4}$$

The mean-square error is

$$\mathcal{E}_{lms} = \sigma_y^2 - \mathbf{c}_{x_y}^{*T} \mathbf{a} = \frac{7}{10} - \begin{bmatrix} \frac{1}{10} & \frac{1}{10} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \frac{3}{5}$$

□